On high-dimensional Lévy-driven Ornstein–Uhlenbeck processes

GESDA Workshop “Non-Linear and High Dimensional Inference”

Claudia Strauch
joint work with Niklas Dexheimer

October 5, 2022

AARHUS UNIVERSITY
DEPARTMENT OF MATHEMATICS

Outline

- **Motivation of high-dimensional models of continuous-time processes**
- **Main results**
- **Heart of the analysis and remaining challenges**

One motivation for high-dimensional models of continuous-time processes

One of the central questions with regard to financial risk management of the entire economy: **How can you build a secure banking system?**

- requires in particular to understand the structure of the interbank lending market in order to be able to predict possible contagions
- according to recent research, interbank lending market has typically a stable structure: there are long-term relationships that are created between some institutions, and most transactions are made through these channels
- structure is not of common knowledge and is hard to observe

One motivation for high-dimensional models of continuous-time processes

One of the central questions with regard to financial risk management of the entire economy: **How can you build a secure banking system?**

- requires in particular to understand the structure of the interbank lending market in order to be able to predict possible contagions
- according to recent research, interbank lending market has typically a stable structure: there are long-term relationships that are created between some institutions, and most transactions are made through these channels
- structure is not of common knowledge and is hard to observe

How to infer the structure of the interbank lending system from the observation of lending activity?

Natural model for interbank lending activity: d-dimensional Ornstein–Uhlenbeck (OU) process fulfilling the stochastic differential equation (SDE)

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \, \mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0,$$

where

- $\mathbf{A}_0 \in \mathbb{R}^{d \times d}$ is a deterministic drift parameter modeling the flow of reserves,
- $W = (W_t)_{t \geq 0}$ is a d-dimensional Brownian motion modeling the random variations of the reserves of a bank.

▹ linear form implies that increased imbalances create increased lending activity

▹ similar models (e.g., with a drift matrix parameter depending on the current state of the reserves in the setting of Feller diffusions) have already been used for that same application

Classical textbook example: Drift parameter estimation for an OU process

- Consider the d-dimensional **OU process** fulfilling

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0.$$

- For 'nice' $\mathbf{A}_0$, it holds

$$X_t = \int_{-\infty}^{t} \exp(-(t-s)\mathbf{A}_0)\,\mathrm{d}W_t \sim \mathcal{N}(0, C_\infty), \text{ with } C_\infty := \int_0^\infty \mathrm{e}^{-s\mathbf{A}_0}\mathrm{e}^{-s\mathbf{A}_0^\top}\,\mathrm{d}s.$$

Classical textbook example: Drift parameter estimation for an OU process

- Consider the d-dimensional **OU process** fulfilling

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0.$$

- For ‘nice’ $\mathbf{A}_0$, it holds

$$X_t = \int_{-\infty}^{t} \exp(-(t-s)\mathbf{A}_0)\,\mathrm{d}W_t \sim \mathcal{N}(0, C_\infty), \text{ with } C_\infty := \int_0^\infty \mathrm{e}^{-s\mathbf{A}_0}\mathrm{e}^{-s\mathbf{A}_0^\top}\,\mathrm{d}s.$$

- How to estimate $\mathbf{A}_0$, given continuous observations $(X_t)_{0\leq t\leq T}$ of the OU process?
- Since the (negative) **log-likelihood function** $\mathcal{L}_T(\cdot)$ is explicitly computed via Girsanov's theorem, it is straightforward to specify the **MLE** $\widehat{\mathbf{A}}_{\mathrm{ML}}$ of $\mathbf{A}_0$ as

$$\widehat{\mathbf{A}}_{\mathrm{ML}} = \operatorname{argmin}_{\mathbf{A}\in\mathbb{R}^{d\times d}} \mathcal{L}_T(\mathbf{A}) = -\left(\int_0^T \mathrm{d}X_t X_t^\top\right)\left(\int_0^T X_t X_t^\top \,\mathrm{d}t\right)^{-1}.$$

For the drift parameter $\mathbf{A}_0$ of the d-dimensional OU process fulfilling the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0,$$

we have the MLE

$$\widehat{\mathbf{A}}_{\mathrm{ML}} = \operatorname{argmin}_{\mathbf{A} \in \mathbb{R}^{d \times d}} \mathcal{L}_T(\mathbf{A}) = -\left(\int_0^T \mathrm{d}X_t X_t^\top\right)\left(\int_0^T X_t X_t^\top \,\mathrm{d}t\right)^{-1}.$$

What happens for ***'large'*** *and* ***'sparse'*** $\mathbf{A}_0$*?*

For the drift parameter $\mathbf{A}_0$ of the d-dimensional OU process fulfilling the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0,$$

we have the MLE

$$\widehat{\mathbf{A}}_{\mathrm{ML}} = \mathrm{argmin}_{\mathbf{A} \in \mathbb{R}^{d \times d}} \mathcal{L}_T(\mathbf{A}) = -\left(\int_0^T \mathrm{d}X_t X_t^\top\right)\left(\int_0^T X_t X_t^\top \,\mathrm{d}t\right)^{-1}.$$

*What happens for **'large'** and **'sparse'*** $\mathbf{A}_0$*?*

- MLE is typically quite inaccurate in high-dimensional setting
- for the real-world application of interbank lending, the interaction structure is sparse

$\hookrightarrow$ Carmona–Fouque model: diagonal (non-diagonal) terms of $\mathbf{A}_0$ correspond to outflows (inflows) of liquidity $\rightarrow$ if each actor receives capital only from a limited number of institutions, this corresponds to row sparsity of $\mathbf{A}_0$

Previous results for high-dimensional OU models

Gaïffas, Matulewicz (2019) [GM19]

- investigate **Lasso** estimators for **classical** OU processes $\mathbf{X}$, assuming that the drift parameter $\mathbf{A}_0$ is ***s*-row-sparse**
- results rely on a **concentration inequality for** $\|X_t\|^2$, which is verified if the drift parameter matrix $\mathbf{A}_0$ is **symmetric**
- rate of convergence:

$$\sqrt{\frac{ds(\log(d) + \log\log T)}{T}}$$

Previous results for high-dimensional OU models

Ciołek, Marushkevych, Podolskij (2020) [CMP20]

- investigate **Lasso** and **Dantzig** estimators for **classical** OU processes $\mathbf{X}$, assuming that the drift parameter $\mathbf{A}_0$ is **s-sparse**, i.e.,

$$\|\mathbf{A}_0\|_0 = \sum_{i,j=1}^{d} 1\{(\mathbf{A}_0)_{ij} \neq 0\} \leq s$$

- concentration inequality verified as soon as $\mathbf{X}$ is **ergodic**
- rate of convergence:

$$\sqrt{\frac{s \log(d^2)}{T}}$$

Rate of convergence?

Consider a d-dimensional (classical) OU model of the form

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0,$$

with underlying observation $(X_t)_{0\leq t\leq T}$. Define the Lasso estimator of $\mathbf{A}_0$ as

$$\widehat{\mathbf{A}}_{\mathrm{lasso}} := \mathrm{argmin}_{\mathbf{A}\in\mathbb{R}^{d\times d}} \left\{\mathcal{L}_T(\mathbf{A}) + \lambda\|\mathbf{A}\|_1\right\}, \quad \lambda > 0.$$

Rate of convergence?

Consider a d-dimensional (classical) OU model of the form

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0,$$

with underlying observation $(X_t)_{0 \leq t \leq T}$. Define the Lasso estimator of $\mathbf{A}_0$ as

$$\widehat{\mathbf{A}}_{\mathrm{lasso}} := \mathrm{argmin}_{\mathbf{A} \in \mathbb{R}^{d \times d}} \left\{ \mathcal{L}_T(\mathbf{A}) + \lambda \|\mathbf{A}\|_1 \right\}, \quad \lambda > 0.$$

Corollary 4.3 in [CMP20]

Fix $\varepsilon \in (0,1)$, and assume that $\|\mathbf{A}_0\|_0 \leq s$. Under some regularity conditions, the Lasso estimator $\widehat{\mathbf{A}}_{\mathrm{lasso}}$ with tuning parameter $\lambda \geq 2\sqrt{\mathrm{const}_1 \frac{\log(2d^2/\varepsilon)}{T}}$ fulfills, if $T \geq T_0 = T_0(\varepsilon, s)$, with probability at least $1 - \varepsilon$,

$$\left\| \widehat{\mathbf{A}}_{\mathrm{lasso}} - \mathbf{A}_0 \right\|_2 \leq \mathrm{const}_2 \sqrt{s} \lambda.$$

Our goals

- Extend results from classical to **Lévy-driven OU processes** $\mathbf{X}$ with background driving Lévy process (BDLP) $\mathbf{Z}$, given as a strong solution of the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geq 0.$$

  - Which are the underlying structures?
  - What are the keys to studying more general high-dimensional models of continuous-time processes?
  - Which are the required (probabilistic) tools?

Our goals

- Extend results from classical to **Lévy-driven OU processes** $\mathbf{X}$ with background driving Lévy process (BDLP) $\mathbf{Z}$, given as a strong solution of the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geq 0.$$

  - Which are the underlying structures?
  - What are the keys to studying more general high-dimensional models of continuous-time processes?
  - Which are the required (probabilistic) tools?
- Suggest regularized estimators of $\mathbf{A}_0$ whose tuning parameters can be chosen **independently of the confidence level**.
- Suggest estimators of $\mathbf{A}_0$ which achieve **minimax optimal rate of convergence**.

Main results

Setting the benchmark: Lower bound

Corollary (Dexheimer & CS (2022))

Let $d \geq 4$ and $s \geq 2d$, and assume that the Lévy triplet of the BDLP $\mathbf{Z}$ is given by $(0, \mathbf{Id}_{d\times d}, 0)$. Then, **for any estimator $\widehat{\mathbf{A}}$**, there **exists some s-sparse matrix $\mathbf{A}_0 \in M_+(\mathbb{R}^d)$** such that, with $\mathbb{P}^{\mathbf{A}_0}$-probability of at least c_0,

$$\|\widehat{\mathbf{A}} - \mathbf{A}_0\|_2 \geq c_1 \sqrt{\frac{s \log(ed^2/s)}{T}}, \qquad \text{for some constants } c_0, c_1 > 0.$$

Notation:

- $M_+(\mathbb{R}^d)$ is the set of all real $d \times d$ matrices such that the real parts of all eigenvalues are positive
- $\mathbb{P}^{\mathbf{A}}$ is the measure induced by the Lévy-driven OU process with drift parameter $\mathbf{A}$, and $\mathbb{P}_t^{\mathbf{A}}$ is the restriction of $\mathbb{P}^{\mathbf{A}}$ on the path space to $\mathcal{F}_t$

Upper bounds: Framework and general assumptions

Consider a d-dimensional **Lévy-driven Ornstein–Uhlenbeck** (OU) process, i.e., a stochastic process $\mathbf{X} = (X_t)_{t \geq 0}$ satisfying the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geq 0, \tag{1}$$

where $\mathbf{Z} = (Z_t)_{t \geq 0}$ is a **Lévy process** with **Lévy triplet** $(b, \mathbf{C} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top, \nu)$.

Upper bounds: Framework and general assumptions

Consider a d-dimensional **Lévy-driven Ornstein–Uhlenbeck** (OU) process, i.e., a stochastic process $\mathbf{X} = (X_t)_{t\geq 0}$ satisfying the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geq 0, \tag{1}$$

where $\mathbf{Z} = (Z_t)_{t\geq 0}$ is a **Lévy process** with **Lévy triplet** $(b, \mathbf{C} = \mathbf{\Sigma}\mathbf{\Sigma}^\top, \nu)$.

Throughout, we impose the following assumptions:

- $\mathbf{A}_0 \in M_+(\mathbb{R}^d)$.
- $\mathbf{C}$ is **strictly positive definite** with minimal (maximal) eigenvalue $\kappa^0_{\min}$ ($\kappa^0_{\max}$).
- The Lévy measure ν admits a **second** moment.
- $\mathbf{X}$ is stationary with invariant distribution μ.
- $\mathbf{A}_0$ is **s-sparse** for $s \in \{d, \dots, d^2\}$.

Lasso and Slope estimators

Let $\mathcal{L}_T(\mathbf{A})$ be the normalized negative **log-likelihood** function, i.e., set

$$\mathcal{L}_T(\mathbf{A}) := -\frac{1}{T}\log\left(\frac{\mathrm{d}\mathbb{P}_T^{\mathbf{A}}}{\mathrm{d}\mathbb{P}_T^{0}}\right), \quad \mathbf{A} \in \mathbb{R}^{d\times d}.$$

The **Lasso** and **Slope** estimators $\widehat{\mathbf{A}}_{\mathrm{lasso}}, \widehat{\mathbf{A}}_{\mathrm{slope}}$ with **tuning parameters** $\lambda_{\mathrm{lasso}}, \lambda_{\mathrm{slope}} \geq 0$ are defined as

$$\widehat{\mathbf{A}}_{\mathrm{lasso}} \in \mathrm{argmin}_{\mathbf{A}\in\mathbb{R}^{d\times d}} \left\{\mathcal{L}_T(\mathbf{A}) + \lambda_{\mathrm{lasso}}\|\mathbf{A}\|_1\right\}, \text{ with } \|\mathbf{A}\|_1 = \sum_{i,j=1}^{d} |\mathbf{A}_{ij}|,$$

$$\widehat{\mathbf{A}}_{\mathrm{slope}} \in \mathrm{argmin}_{\mathbf{A}\in\mathbb{R}^{d\times d}} \left\{\mathcal{L}_T(\mathbf{A}) + \lambda_{\mathrm{slope}}\|\mathbf{A}\|_*\right\}, \text{ with } \|\mathbf{A}\|_* = \sum_{i=1}^{d^2} \mathbf{A}_i^{\#}\sqrt{\log\left(\frac{2d^2}{i}\right)},$$

where $\mathbf{A}^{\#} \in \mathbb{R}^{d^2}$ is a **non-increasing rearrangement** of the absolute values of $\mathbf{A}$'s entries.

SLOPE—ADAPTIVE VARIABLE SELECTION VIA CONVEX OPTIMIZATION

BY MAŁGORZATA BOGDAN[*,1], EWOUT VAN DEN BERG[†,2], CHIARA SABATTI[‡,3], WEIJIE SU[‡,4] AND EMMANUEL J. CANDÈS[‡,5]

Wrocław University of Technology[*], *IBM T.J. Watson Research Center*[†] *and Stanford University*[‡]

We introduce a new estimator for the vector of coefficients β in the linear model $y = X\beta + z$, where X has dimensions $n \times p$ with p possibly larger than n. SLOPE, short for Sorted L-One Penalized Estimation, is the solution to

$$\min_{b \in \mathbb{R}^p} \frac{1}{2}\|y - Xb\|_{\ell_2}^2 + \lambda_1 |b|_{(1)} + \lambda_2 |b|_{(2)} + \cdots + \lambda_p |b|_{(p)},$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and $|b|_{(1)} \geq |b|_{(2)} \geq \cdots \geq |b|_{(p)}$ are the decreasing absolute values of the entries of b. This is a convex program and we demonstrate a solution algorithm whose computational complexity is roughly comparable to that of classical ℓ_1 procedures such as the Lasso. Here, the

Upper bounds: Starting point of the investigations

Lemma (Dexheimer & CS (2022))

Let $h\colon \mathbb{R}^{d\times d} \longrightarrow \mathbb{R}$ be a **convex** function. If $\widehat{\mathbf{A}}$ is a solution of the **minimization problem** $\min_{\mathbf{A}\in\mathbb{R}^{d\times d}} \{\mathcal{L}_T(\mathbf{A}) + h(\mathbf{A})\}$, then $\widehat{\mathbf{A}}$ satisfies for all $\mathbf{A} \in \mathbb{R}^{d\times d}$

$$\begin{aligned}\left\|(\widehat{\mathbf{A}} - \mathbf{A}_0)X\right\|_{L^2}^2 - \frac{\kappa^0_{\max}}{\kappa^0_{\min}}\left\|(\mathbf{A} - \mathbf{A}_0)X\right\|_{L^2}^2 \\ \le 2\kappa^0_{\max}\left(\langle \boldsymbol{\epsilon}_T, \boldsymbol{\Sigma}^{-1}(\mathbf{A} - \widehat{\mathbf{A}})\rangle_2 + h(\mathbf{A}) - h(\widehat{\mathbf{A}})\right) - \left\|(\widehat{\mathbf{A}} - \mathbf{A})X\right\|_{L^2}^2,\end{aligned}$$

where

$$\|\mathbf{A}X\|_{L^2}^2 = \frac{1}{T}\int_0^T \|\mathbf{A}X_s\|^2 \,\mathrm{d}s \quad \textit{and} \quad \boldsymbol{\epsilon}_T^\top := \frac{1}{T}\int_0^T X_s \,\mathrm{d}W_s^\top,$$

with $(W_s)_{s\ge 0}$ being a $\mathbb{P}^{\mathbf{A}_0}$-Wiener process.

Proof of our main results is based on two central elements:

(1) We confine ourselves to the **study of a benign event**, in our context of the form

$$\mathcal{E} := \left\{ \inf_{\mathbf{B} \in \mathbb{R}^{d \times d} \setminus \{0\}} \frac{\|\mathbf{B}X\|_{L^2}^2}{\|\mathbf{B}\|_2^2} > c_1 \right\} \bigcap \left\{ \sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \mathbf{B} \neq 0} \frac{\langle \boldsymbol{\epsilon}_T, \mathbf{B} \rangle_2}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c_2}{\sqrt{T}} \right\}.$$

- investigation **on this event** is driven by **purely deterministic arguments**, which can be developed analogously to the high-dimensional linear regression model

Proof of our main results is based on two central elements:

(1) We confine ourselves to the **study of a benign event**, in our context of the form

$$\mathcal{E} := \left\{ \inf_{\mathbf{B} \in \mathbb{R}^{d \times d} \setminus \{0\}} \frac{\|\mathbf{B}X\|_{L^2}^2}{\|\mathbf{B}\|_2^2} > c_1 \right\} \bigcap \left\{ \sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \mathbf{B} \neq 0} \frac{\langle \boldsymbol{\epsilon}_T, \mathbf{B} \rangle_2}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c_2}{\sqrt{T}} \right\}.$$

- investigation **on this event** is driven by **purely deterministic arguments**, which can be developed analogously to the high-dimensional linear regression model

(2) We show that the **event $\mathcal{E}$ is of high probability**.

Proof of our main results is based on two central elements:

(1) We confine ourselves to the **study of a benign event**, in our context of the form

$$\mathcal{E} := \left\{ \inf_{\mathbf{B}\in\mathbb{R}^{d\times d}\setminus\{0\}} \frac{\|\mathbf{B}X\|_{L^2}^2}{\|\mathbf{B}\|_2^2} > c_1 \right\} \bigcap \left\{ \sup_{\mathbf{B}\in\mathbb{R}^{d\times d}:\mathbf{B}\neq 0} \frac{\langle \boldsymbol{\epsilon}_T, \mathbf{B}\rangle_2}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c_2}{\sqrt{T}} \right\}.$$

- investigation **on this event** is driven by **purely deterministic arguments**, which can be developed analogously to the high-dimensional linear regression model

(2) We show that the **event $\mathcal{E}$ is of high probability**.

Analysis of the first sub-event?

- verify a property of **restricted eigenvalue-type**
- similarly to the Gaussian case, a **concentration condition** can be used to show this

Fundamental concentration condition: Key to the proofs

($\mathcal{H}$) There exists a function $H \colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that

- for any $T, r > 0$, the functions $H(T, \cdot)$ and $H(\cdot, r)$ are non-increasing and such that

$$\forall r > 0, \quad \lim_{T \to \infty} H(T, r) = 0;$$

- for any vector $u \in \mathbb{R}^d$ with $\|u\| \leq 1$, it holds

$$\forall T, r > 0, \quad \mathbb{P}\left(|u^\top(\widehat{\mathbf{C}}_T - \mathbf{C}_\infty)u| > r\right) \leq H(T, r),$$

where

$$\widehat{\mathbf{C}}_T := \frac{1}{T}\int_0^T X_s X_s^\top \,\mathrm{d}s \quad \text{and} \quad \mathbf{C}_\infty := \int x x^\top \mu(\mathrm{d}x).$$

Proof of our main results is based on two central elements:

(1) We confine ourselves to the study of a benign event, in our context of the form

$$\mathcal{E} := \left\{ \inf_{\mathbf{B}\in\mathbb{R}^{d\times d}\setminus\{0\}} \frac{\|\mathbf{B}X\|_{L^2}^2}{\|\mathbf{B}\|_2^2} > c_1 \right\} \bigcap \left\{ \sup_{\mathbf{B}\in\mathbb{R}^{d\times d}:\mathbf{B}\neq 0} \frac{\langle \boldsymbol{\epsilon}_T, \mathbf{B}\rangle_2}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c_2}{\sqrt{T}} \right\}.$$

(2) We show that the event $\mathcal{E}$ **is of high probability**.

Proof of our main results is based on two central elements:

(1) We confine ourselves to the study of a benign event, in our context of the form

$$\mathcal{E} := \left\{ \inf_{\mathbf{B} \in \mathbb{R}^{d \times d} \setminus \{0\}} \frac{\|\mathbf{B}X\|_{L^2}^2}{\|\mathbf{B}\|_2^2} > c_1 \right\} \bigcap \left\{ \sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \mathbf{B} \neq 0} \frac{\langle \boldsymbol{\epsilon}_T, \mathbf{B} \rangle_2}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c_2}{\sqrt{T}} \right\}.$$

(2) We show that the event **$\mathcal{E}$ is of high probability**.

Analysis of the second sub-event?

- involves both the process $\boldsymbol{\epsilon}_T$ and the norm

$$\|\mathbf{B}\|_{\mathrm{S}} := \|\mathbf{B}\|_* \vee \sqrt{\log(4/\varepsilon_0)}\|\mathbf{B}\|_2$$

- main differences with previous studies of high-dimensional OU models do *not* arise because of the structure of the process, but because of the modified statistical approach
- controlling the second sub-event provides the **key to obtain the optimal** convergence rate

Deviation inequality for the stochastic error

- Introduce the event

$$Q_T := \left\{ \sup_{u \in \mathbb{R}^d : \ \|u\| \leq 1} |u^\top (\widehat{\mathbf{C}}_T - \mathbf{C}_\infty) u| \leq \frac{\kappa_{\min}}{2} \right\}.$$

- If assumption $(\mathcal{H})$ holds, we have

$$\mathbb{P}(Q_T) \geq 1 - (21(d \wedge \mathrm{e}))^d \, H\left(T, \frac{\kappa_{\min}}{6}\right).$$

Deviation inequality for the stochastic error

- Introduce the event

$$Q_T := \left\{ \sup_{u \in \mathbb{R}^d : \|u\| \leq 1} |u^\top (\widehat{\mathbf{C}}_T - \mathbf{C}_\infty) u| \leq \frac{\kappa_{\min}}{2} \right\}.$$

- If assumption $(\mathcal{H})$ holds, we have

$$\mathbb{P}\left(Q_T\right) \geq 1 - (21(d \wedge \mathrm{e}))^d H\left(T, \frac{\kappa_{\min}}{6}\right).$$

- Furthermore, if assumption $(\mathcal{H})$ holds, we control for any $\mathbf{B} \in \mathbb{R}^{d \times d}$ the probability that

$$\frac{1}{T} \underbrace{\int_0^T \|\mathbf{B} X_s\|_2^2 \, \mathrm{d}s}_{\text{quadratic variation of the stochastic error } T\langle \boldsymbol{\epsilon}_T, \mathbf{B} \rangle_2} > \text{const.} \|\mathbf{B}\|_2^2.$$

Deviation inequality for the stochastic error

- By **Bernstein's inequality for continuous martingales**, we get for any $u > 0$

$$\mathbb{P}\left(|\langle \epsilon_T, \mathbf{B}\rangle \, 1_{Q_T}| \geq \sqrt{\frac{2c_2}{T}} u \|\mathbf{B}\|_2\right) \leq \mathbb{P}\left(\Big|\int_0^T (\mathbf{B}X_s)^\top \, \mathrm{d}W_s\Big| \geq \sqrt{2c_2 T} u \|\mathbf{B}\|_2, \int_0^T \|\mathbf{B}X_s\|_2^2 \, \mathrm{d}s \leq 2c_2 T \|\mathbf{B}\|_2^2\right) \leq 2\mathrm{e}^{-u^2}.$$

Deviation inequality for the stochastic error

- By **Bernstein's inequality for continuous martingales**, we get for any $u > 0$

$$\mathbb{P}\left(|\langle \epsilon_T, \mathbf{B}\rangle\ 1_{Q_T}| \geq \sqrt{\frac{2c_2}{T}} u \|\mathbf{B}\|_2\right)$$
$$\leq \mathbb{P}\left(\Big|\int_0^T (\mathbf{B}X_s)^\top \,\mathrm{d}W_s\Big| \geq \sqrt{2c_2T} u \|\mathbf{B}\|_2, \int_0^T \|\mathbf{B}X_s\|_2^2 \,\mathrm{d}s \leq 2c_2T\|\mathbf{B}\|_2^2\right)$$
$$\leq 2\mathrm{e}^{-u^2}.$$

- **Generic chaining** device then implies, for some constant $c > 0$,

$$\mathbb{P}\left(\sup_{\mathbf{B}\in\mathbb{R}^{d\times d}\setminus\{0\}} \frac{\langle \epsilon_T, \mathbf{B}\rangle\ 1_{Q_T}}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c}{\sqrt{T}}\left(w(\mathcal{D}^*) + u\,\mathrm{rad}(\mathcal{D}^*)\right)\right) \geq 1 - 2\mathrm{e}^{-u^2}, \quad u > 0,$$

where $w(\mathcal{D}^*)$ and $\mathrm{rad}(\mathcal{D}^*)$ are the **Gaussian width** respectively **radius** of

$$\mathcal{D}^* = \left\{\mathbf{B} \in \mathbb{R}^{d\times d} : \|\mathbf{B}\|_{\mathrm{S}} = 1\right\}.$$

Deviation inequality for the stochastic error

- For the **Gaussian width**, we have with $Z \sim \mathcal{N}(0, \mathbf{Id}_{d^2 \times d^2})$

$$w(\mathcal{D}^*) = \mathbb{E}\left[\sup_{\mathbf{B} \in \mathcal{D}^*} \langle \mathrm{vec}(\mathbf{B}), Z \rangle\right] \leq \mathbb{E}\left[\sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \|\mathbf{B}\|_* \leq 1} \langle \mathrm{vec}(\mathbf{B}), Z \rangle\right] = \mathbb{E}\left[\sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \|\mathbf{B}\|_* \leq 1} \sum_{i=1}^{d^2} \mathbf{B}_i^{\#} Z_i^{\#}\right]$$

$$\leq \mathbb{E}\left[\max_{i=1,\ldots,d^2} Z_i^{\#} \left(\log\left(\frac{2d^2}{i}\right)\right)^{-1/2}\right] \leq \sqrt{\frac{\pi}{4\log(2)}} + 4.$$

Deviation inequality for the stochastic error

- For the **Gaussian width**, we have with $Z \sim \mathcal{N}(0, \mathbf{Id}_{d^2 \times d^2})$

$$w(\mathcal{D}^*) = \mathbb{E}\left[\sup_{\mathbf{B} \in \mathcal{D}^*} \langle \mathrm{vec}(\mathbf{B}), Z \rangle\right] \leq \mathbb{E}\left[\sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \|\mathbf{B}\|_* \leq 1} \langle \mathrm{vec}(\mathbf{B}), Z \rangle\right] = \mathbb{E}\left[\sup_{\mathbf{B} \in \mathbb{R}^{d \times d} : \|\mathbf{B}\|_* \leq 1} \sum_{i=1}^{d^2} \mathbf{B}_i^{\#} Z_i^{\#}\right]$$

$$\leq \mathbb{E}\left[\max_{i=1,\ldots,d^2} Z_i^{\#} \left(\log\left(\frac{2d^2}{i}\right)\right)^{-1/2}\right] \leq \sqrt{\frac{\pi}{4\log(2)}} + 4.$$

- Combining with $\mathrm{rad}(\mathcal{D}^*) \leq \sqrt{\log(4/\varepsilon_0)}^{-1}$, we obtain for a constant $c_* > 0$

$$\mathbb{P}\left(\sup_{\mathbf{B} \in \mathbb{R}^{d \times d} \setminus \{0\}} \frac{\langle \epsilon_T, \mathbf{B} \rangle \, 1_{Q_T}}{\|\mathbf{B}\|_{\mathrm{S}}} \leq \frac{c_*}{\sqrt{T}}\right) \geq 1 - \frac{\varepsilon_0}{2},$$

thus showing the **desired deviation inequality**.

Main result: Lasso

Corollary (of the main theorem) (Dexheimer & CS (2022))

Grant Assumption $(\mathcal{H})$, fix $\varepsilon_0 \in (0,1)$, and let $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}_{\mathrm{lasso}}$ be the **Lasso estimator** with **tuning parameter**

$$\lambda_T \geq 2c_* \sqrt{\frac{\kappa^*}{T} \log\left(\frac{2\mathrm{e}d^2}{s}\right)}.$$

Then, for

$$T > T_0(\varepsilon_0) := \inf\left\{T > 0 : (21(d \wedge \mathrm{e}))^d H\left(T, \frac{\kappa_{\min}}{6}\right) \leq \frac{\varepsilon_0}{2}\right\}, \tag{2}$$

with probability larger than $1 - \frac{1}{2}(\varepsilon_0 + \varepsilon_1)$, for all $\varepsilon_1 \in (0,1)$:

$$\|\widehat{\mathbf{A}} - \mathbf{A}_0\|_2^2 \leq \frac{16s\left(\kappa_{\max}^0 \lambda_T\right)^2}{\kappa_{\min}^2} \left(1 \vee \frac{\log(4\varepsilon_1^{-1})}{s\log(2\mathrm{e}d^2/s)}\right).$$

Main result: Slope

Corollary (Dexheimer & CS (2022))

Grant Assumption ($\mathcal{H}$), fix $\varepsilon_0 \in (0,1)$, and let $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}_{\text{slope}}$ be the **Slope estimator** with **tuning parameter**

$$\lambda_T := 2c_S T^{-1/2}.$$

Then, for $T > T_0(\varepsilon_0)$ and $T_0(\cdot)$ defined as in (2), with probability larger than $1 - \frac{1}{2}(\varepsilon_0 + \varepsilon_1)$, for all $\varepsilon_1 \in (0,1)$:

$$\|\widehat{\mathbf{A}} - \mathbf{A}_0\|_2^2 \leq 64\left(\kappa_{\max}^0 c_S\right)^2 \frac{s\log\left(2ed^2/s\right)}{T\kappa_{\min}^2}\left(1 \vee \frac{\log(4\varepsilon_1^{-1})}{s\log(2ed^2/s)}\right).$$

Heart of the analysis and remaining challenges

We have identified the fundamental concentration condition:

($\mathcal{H}$) There exists a function $H\colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that

- for any $T, r > 0$, the functions $H(T, \cdot)$ and $H(\cdot, r)$ are non-increasing and such that

$$\forall r > 0, \quad \lim_{T\to\infty} H(T, r) = 0;$$

- for any vector $u \in \mathbb{R}^d$ with $\|u\| \le 1$, it holds

$$\forall T, r > 0, \quad \mathbb{P}\left(|u^\top(\widehat{\mathbf{C}}_T - \mathbf{C}_\infty)u| > r\right) \le H(T, r),$$

where

$$\widehat{\mathbf{C}}_T := \frac{1}{T}\int_0^T X_s X_s^\top \,\mathrm{d}s \quad \text{and} \quad \mathbf{C}_\infty := \int x x^\top \mu(\,\mathrm{d}x).$$

We have identified the fundamental concentration condition:

($\mathcal{H}$) There exists a function $H\colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that

- for any $T, r > 0$, the functions $H(T,\cdot)$ and $H(\cdot, r)$ are non-increasing and such that

$$\forall r > 0, \quad \lim_{T\to\infty} H(T,r) = 0;$$

- for any vector $u \in \mathbb{R}^d$ with $\|u\| \leq 1$, it holds

$$\forall T, r > 0, \quad \mathbb{P}\left(|u^\top(\widehat{\mathbf{C}}_T - \mathbf{C}_\infty)u| > r\right) \leq H(T,r),$$

where

$$\widehat{\mathbf{C}}_T := \frac{1}{T}\int_0^T X_s X_s^\top \,\mathrm{d}s \quad \text{and} \quad \mathbf{C}_\infty := \int x x^\top \mu(\mathrm{d}x).$$

Which assumptions on the underlying process X are needed to verify this?

Some known concentration results for Markov processes

Let $\mathbf{X}$ be a nice ergodic Markov processes with semigroup $(P_t)_{t\geq 0}$, invariant distribution μ, and denote

$$C_\nu(f,T,x) := \mathbb{P}^\nu\left(\left|\frac{1}{T}\int_0^T f(X_s)\,\mathrm{d}s - \mu(f)\right| > x\right), \quad f \in L^2(\mu), x, T > 0.$$

Bounds on $C_\nu(f,T,x)$ have been mostly studied with two approaches (Lyapunov vs. Poincaré [BCG08]):

1. **functional inequalities**, in particular **Poincaré inequality** and **log-Sobolev inequality**,
2. **mixing assumptions**, e.g.,

$$\alpha_\nu(t) := \sup_{s\geq 0} \sup_{A\in\sigma(X_u, u\leq s), B\in\sigma(X_u, u\geq s+t)} |\mathbb{P}^\nu(A\cap B) - \mathbb{P}^\nu(A)\mathbb{P}^\nu(B)| \leq \varphi(t) \to_{t\to\infty} 0$$

Some known concentration results for Markov processes

- [Lez01]: Suppose $\nu \ll \mu$, $\mathrm{d}\nu/\mathrm{d}\mu \in L^2(\mu)$ and $\|f\|_\infty < \infty$. If μ satisfies the **Poincaré inequality**, then we have the **Bernstein inequality**

$$C_\nu(f,T,x) \leq 2 \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|_{L^2(\mu)} \exp\left(- \frac{T x^2}{2\left(\sigma^2(f) + 2C\|f\|_\infty x\right)} \right),$$

where $\sigma^2(f) = \lim_{t\to\infty} t^{-1} \operatorname{Var}_{\mathbb{P}^\mu}\left(\int_0^t f(X_s)\,\mathrm{d}s\right)$.

Some known concentration results for Markov processes

- [Lez01]: Suppose $\nu \ll \mu$, $\mathrm{d}\nu / \mathrm{d}\mu \in L^2(\mu)$ and $\|f\|_\infty < \infty$. If μ satisfies the **Poincaré inequality**, then we have the **Bernstein inequality**

$$C_\nu(f, T, x) \leq 2 \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|_{L^2(\mu)} \exp\left(-\frac{T x^2}{2\left(\sigma^2(f) + 2C\|f\|_\infty x\right)} \right),$$

  where $\sigma^2(f) = \lim_{t\to\infty} t^{-1} \operatorname{Var}_{\mathbb{P}^\mu}\left(\int_0^t f(X_s)\,\mathrm{d}s \right)$.

- [GGW14]: If μ satisfies the **log-Sobolev inequality**, $\mu(f) = 0$ and $\mu(\exp(\lambda_\pm f^\pm)) < \infty$ for some $\lambda_\pm > 0$, then we have the **Bernstein inequality**

$$C_\nu(f, T, x) \leq 2 \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|_{L^2(\mu)} \exp\left(-\frac{T x^2}{2\left(\sigma^2(f) + C_P(\Lambda^*)^{-1}(2C_{LS}/C_P)x\right)} \right),$$

  where $\Lambda_\pm^*$ is the Legendre transform of $[0, \lambda_\pm] \ni s \mapsto \Lambda_\pm(s) := \log \mu(\exp(s(\pm f)))$ and $\Lambda^* = \Lambda_+^* \vee \Lambda_-^*$.

Application to the high-dimensional anaysis of the classical OU model: Sufficient conditions for assumption ($\mathcal{H}$) in the Gaussian case

- [GM19]: ($\mathcal{H}$) holds under assumptions implied by ergodicity assumption if $\mathbf{A}_0$ is symmetric
- result was achieved by exploiting that symmetricity of $\mathbf{A}_0$ implies μ to fulfill a log-**Sobolev inequality**

Application to the high-dimensional anaysis of the classical OU model: Sufficient conditions for assumption ($\mathcal{H}$) in the Gaussian case

- [GM19]: ($\mathcal{H}$) holds under assumptions implied by ergodicity assumption if $\mathbf{A}_0$ is symmetric
- result was achieved by exploiting that symmetricity of $\mathbf{A}_0$ implies μ to fulfill a **log-Sobolev inequality**
- [CMP20] extended this finding to the general case of possibly non-symmetric $\mathbf{A}_0$, i.e., $\mathbf{A}_0 \in M_+(\mathbb{R}^d)$ already implies ($\mathcal{H}$) in the classical Gaussian case
- proof of this result relies on **Malliavin calculus methods**
- in both papers, the function H in ($\mathcal{H}$) is of the form

$$H(T, r) = 2\exp(-T H_0(r)), \quad T, r > 0,$$

where H_0 is positive and increasing

Sufficient conditions for assumption $(\mathcal{H})$ in the Lévy-driven case

- achieving similar results as in the Gaussian setting is a challenging task
- application of the stochastic Fubini theorem, combined with classical martingale results, yields that $(\mathcal{H})$ is fulfilled as soon as the Lévy measure ν of the BDLP admits a fourth moment

Sufficient conditions for assumption $(\mathcal{H})$ in the Lévy-driven case

- achieving similar results as in the Gaussian setting is a challenging task
- application of the stochastic Fubini theorem, combined with classical martingale results, yields that $(\mathcal{H})$ is fulfilled as soon as the Lévy measure ν of the BDLP admits a fourth moment

Proposition (Dexheimer & CS (2022))

Assume that ν admits a fourth moment. Then, there exists a constant $c > 0$ such that, for all $u \in \mathbb{R}^d$ fulfilling $\|u\| \leq 1$,

$$\mathbb{P}\left(|u^\top(\widehat{\mathbf{C}}_T - \mathbf{C}_\infty)u| \geq r\right) \leq \frac{c}{t(r \wedge r^2)} + \frac{c}{(tr)^2}.$$

In particular, Assumption $(\mathcal{H})$ is fulfilled.

The central challenge

- concentration result in the Lévy case is obviously weaker than its Gaussian counterpart in the sense of the temporal decay not being exponential but polynomial
- primary influence of this is on the value of the threshold value T_0 specified which increases
- our main results are developed in such a way that they only rely on Assumption ($\mathcal{H}$) in its general form, and so it would be easy to implement results implying an exponential decay in the Lévy-driven case to achieve values of T_0 similar to the Gaussian case

The central challenge

- concentration result in the Lévy case is obviously weaker than its Gaussian counterpart in the sense of the temporal decay not being exponential but polynomial
- primary influence of this is on the value of the threshold value T_0 specified which increases
- our main results are developed in such a way that they only rely on Assumption $(\mathcal{H})$ in its general form, and so it would be easy to implement results implying an exponential decay in the Lévy-driven case to achieve values of T_0 similar to the Gaussian case
- **note:** results of [CMP20] in the Gaussian OU case require that

$$T \geq T_0 = \text{const.}\, s \log d$$

Summary

- We have extended the results from classical to **Lévy-driven OU processes** $\mathbf{X}$ with background driving Lévy process (BDLP) $\mathbf{Z}$, given as a strong solution of the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geq 0.$$

  - Concentration condition is the key to the in-depth high-dimensional study!
- We suggest regularized estimators of $\mathbf{A}_0$ whose tuning parameters can be chosen **independently of the confidence level**.
- We show that the suggested estimators of $\mathbf{A}_0$ achieve the **minimax optimal rate of convergence**.

Summary

- We have extended the results from classical to **Lévy-driven OU processes** $\mathbf{X}$ with background driving Lévy process (BDLP) $\mathbf{Z}$, given as a strong solution of the SDE

$$\mathrm{d}X_t = -\mathbf{A}_0 X_t \,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geq 0.$$

  - Concentration condition is the key to the in-depth high-dimensional study!
- We suggest regularized estimators of $\mathbf{A}_0$ whose tuning parameters can be chosen **independently of the confidence level**.
- We show that the suggested estimators of $\mathbf{A}_0$ achieve the **minimax optimal rate of convergence**.

Thank you!